

DOCUMENT RESUME

ED 075 492

TM 002 571

AUTHOR Jongsma, Eugene A.
TITLE Viewing Standardized Social Studies Achievement Tests
from a Reading Perspective.
PUB DATE Nov 72
NOTE 17p.; Paper presented at Annual Meeting of Mid-South
Educational Research Association (New Orleans,
Louisiana, November 1972)
EDRS PRICE MF-\$0.65 HC-\$3.29
DESCRIPTORS *Achievement Tests; Psychometrics; *Reading Skills;
*Social Studies; *Standardized Tests; Test
Interpretation; *Test Validity

ABSTRACT

A critical examination is made of standardized social studies achievement tests from a psychometric and reading perspective. Five major issues are identified that detract from the meaningful interpretation of student performance on standardized social studies tests. The issues discussed are (1) the reading dependency of social studies items, (2) the picture-dependency of social studies items, (3) the cognitive skills assessed by social studies items, (4) the lack of an adequate system of item development, and (5) the lack of content validity. Each of the issues is defined and related to test validity and interpretation. The central theme of the study is the inadequacy of the content validity of most standardized social studies tests. Unless test publishers specify more explicitly the elements of content and types of behavior sampled on their tests, test users will have great difficulty in making meaningful interpretations of student performance.
(Author/DB)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIG-
INATING IT. POINTS OF VIEW OR OPIN-
IONS STATED DO NOT NECESSARILY
REPRESENT OFFICIAL OFFICE OF EDU-
CATION POSITION OR POLICY

VIEWING STANDARDIZED SOCIAL STUDIES ACHIEVEMENT TESTS
FROM A READING PERSPECTIVE

Eugene A. Jongoma
Louisiana State University in New Orleans

Paper presented at the Annual Meeting of the Mid-South
Educational Research Association
New Orleans, November 1972

ED 075492

TM 002 571

VIEWS OF STANDARDIZED SOCIAL STUDIES ACHIEVEMENT TESTS
FROM A READING PERSPECTIVE

Eugene A. Jongsma
Louisiana State University in New Orleans

It comes as no surprise to say that a close relationship exists between reading ability and achievement in the social studies. Virtually all standardized social studies tests involve some reading. There are some exceptions to this such as Preston and Duffey's Primary Social Studies Test (1967) which uses pictures but, by and large, reading is required on nearly all standardized social studies tests.

How important is this relationship? To what extent is performance on a standardized social studies test influenced by reading ability and general test-taking skills? We (Gaines and Jongsma, 1972) recently conducted a study in which we attempted to raise student performance on a standardized achievement test by teaching the students a few basic reading and test-taking skills. The students were lower socioeconomic fifth graders who we assumed were not "test-wise". We developed an instructional package called "Test-Taking Tips" which consisted of illustrations and exercises for the students to work. The unit covered five major topics--(1) motivation, (2) following directions, (3) guessing, (4) reading comprehension, and (5) test behavior. A random selection of students worked through the unit in approximately one hour the day before the standardized test was administered. Results showed that students who worked through the unit made significantly higher scores on several of the subtests than their counterparts who had not seen the unit. The

social studies section was one of the subtests on which significant differences were found. That is, we improved the students' social studies achievement, not by teaching them anything about social studies, but by alerting them to a few reading and test-taking skills. Another interesting sidelight to this study was the unusually high correlation of .83 which we found between total reading score and the social studies subtest. However, similarly high correlations between reading achievement and social studies achievement have been reported by other researchers (Thomas, 1967; Wash, 1968; Gaines, 1971). This suggests the influence reading ability has on performance on social studies tests.

The purpose of this paper is to present a critical analysis of current standardized social studies achievement tests. The analysis will be based on psychometric considerations and the influence of related reading skills. Five major issues will be identified and discussed.

1. Reading-Dependency of Social Studies Test Items

One of the formats that is commonly found on standardized social studies tests is the procedure whereby the student is given a passage to read followed by multiple-choice comprehension questions. This format looks no different than conventional reading comprehension tests except for the social studies content. The assumption is that the multiple-choice items are based directly on the passage and that the student must comprehend the passage in order to correctly answer the items. That is, if the items are reading-dependent, students will not

be able to obtain better than a chance score without having read the prerequisite passage. Thus, reading-dependency refers to the relationship between multiple-choice test items and the passage on which they are based.

Several studies have shown that such is not the case. Preston (1964) found that the mean score of a group of students not reading the passage was significantly greater than chance on the Reading Comprehension section of the Cooperative English Test. Weaver and Bickley (1967) randomly selected items from several standardized reading comprehension tests listed in Buros' Sixth Mental Measurement Yearbook and administered them to college students. Subjects answering without the passages answered 67 percent as many items as the subjects who had access to the passages. Mitchell (1967) obtained similar results with fourth graders using the Gates Basic Reading Test. Tuinman (1970) found a lack of reading-dependency for items on the STEP Reading Test. While all these studies have used reading comprehension tests, I suspect we would find similar results with standardized social studies tests that use the same format.

Perhaps the following sample items will help to illustrate the point:

Passage*
Comparison of transatlantic travel
between Pilgrims and travelers of 1955

25. Which is the best reason why the 1955 travelers were more comfortable than the Pilgrims?
- A. The Pilgrims were poor.
 - B. The 1955 travelers were more intelligent.
 - C. The ocean was less stormy.
 - D. Between 1620 and 1955 many inventions had been made.
26. What kind of power was used to move the Mayflower?
- A. Wind blowing on sails
 - B. Many oars pulled by slaves
 - C. Steam engines using coal
 - D. Gasoline engines

Passage*
Description of a family camping
and cooking out

16. Which is the best reason why John's family should observe fire laws in forests?
- E. Those who disobey are punished.
 - F. Animals are frightened by campfires.
 - G. Fires can cause great damage in forests.
 - H. Few people know how to build a safe fire.
17. Who makes the laws about fires in forests?
- A. The forest rangers
 - B. The government
 - C. The people who sell the timber
 - D. The men who cut the timber

*Sequential Test of Educational Progress, Social Studies,
Form 4A, Educational Testing Service, 1956

Each of these items was preceded by a passage. Students were to answer the items based upon information gained from the passage. However, as one can see many students would be able to answer such items

without reading the relevant passage at all. In short, these items do not appear to be reading-dependent.

Some may ask if it's indeed necessary for social studies items to be reading-dependent. The answer to that question is a matter of test validity and interpretation. If the purpose in testing is to assess whether the student has achieved some predetermined level of comprehension of social studies content, then reading-dependency is not important. In this case, the sources of information from which the student draws when answering are not important. On the other hand, if the purpose is to determine how much information a student is able to gain from reading social studies material, the items must be reading dependent. It is doubtful if test publishers have come to grips with this issue.

Reading-dependency is a relative matter that is related to the knowledge and experiential background of the student. For one student an item may be reading-dependent while for a more sophisticated student, the same item is not reading-dependent. For practical reasons judgments regarding the degree of reading-dependency cannot be made in terms of individuals but have to be assessed on a group basis.

2. Picture-Dependency of Social Studies Test Items

Much of what was said in the previous section could also apply to the use of pictures on standardized social studies tests. On some tests students are asked questions which are supposedly based upon pictures included in the test. Careful examination of such questions, however, reveals that many of them could probably be answered without

even referring to the picture. The following examples were taken from standardized social studies tests that are currently on the market:

Picture of the produce section
of a grocery store in Ohio*

1. Which food was raised on a farm somewhere in our country?
 - A. Tea
 - B. Apples
 - C. Cocoa
 - D. Coffee

4. What food most likely traveled part of the way to the store by boat?
 - E. Oranges
 - F. Peas
 - G. Bananas
 - H. Celery

5. Which food most likely traveled farthest to reach the store?
 - A. Pineapples
 - B. Lettuce
 - C. Eggs
 - D. Peaches

Series of pictures which tell the
story of bread-making*

20. Which picture should come first?
 - E. The grain elevator
 - F. The flour mill
 - G. The bakery
 - H. The shocks of wheat

21. Which picture should come next after the grain elevator?
 - A. The bread on a store shelf
 - B. The flour mill
 - C. The bakery
 - D. The slice of bread being buttered

Most students would be able to answer these items without referring to the pictures provided in the test. What purpose do the pictures serve in this case? My only conclusion is that they may serve as an aid to students weak in word identification skills. For example, the student who is unable to identify the word "pineapples" may refer to the pictures, recognize the visual referent of "pineapples", and proceed to answer the item. I seriously question if this was the purpose the publishers intended the pictures to serve.

Once again the issue is one of test validity. If the intent is to assess the student's ability to recognize certain social studies concepts in pictures then the items must be picture-dependent. Perhaps what is needed is some rethinking of the role pictures should play on social studies test.

3. Cognitive Skills Assessed by Social Studies Items

If one critically examines the items found on standardized social studies tests, he should find that in some cases the items strongly resemble the kinds of items often found on group intelligence tests. For example, many standardized social studies tests include an assessment of the student's social studies vocabulary. Although this would appear appropriate, some of the vocabulary sections of social studies

tests seem to extend into other cognitive areas as well. Perhaps the following examples taken from the Metropolitan Achievement Test will illustrate this point. The directions to the students are: "Read each set of headings and the list of items following each set. Each item is most closely associated with, or fits best under, one of the headings. Decide which heading is best for each item."

Selected Items from the Intermediate Level*

<u>Headings</u>	<u>Items</u>
A. Commerce and Trade	27. minority rights
B. Communication	28. profit
C. Government	35. A-bomb
D. Inventions	40. power loom

Selected Items from the Advanced Level**

<u>Headings</u>	<u>Items</u>
A. Authors and Journalists	21. Babe Ruth
B. Educational, Religious and Social Reformers	24. Eli Whitney
C. Scientists and Inventors	25. Edgar A. Poe
D. Leaders in Entertainment, Sports, and Theater	32. Wilt Chamberlain

* Metropolitan Achievement Test, Social Studies, Intermediate Level, Form F, Harcourt Brace and Jovanovich, 1970

**Metropolitan Achievement Test, Social Studies, Advanced Level, Form F, Harcourt Brace and Jovanovich, 1970

Critical examination of these items reveals several important points. First, it appears that these items require the student to employ some sort of classification skill over and above the knowledge of the vocabulary used. Second, the vocabulary items are presented in isolation which is artificial and not reflective of the contextual settings used in instruction. This is especially misleading for vocabulary terms or phrases which may have a wide range of connotative

and denotative meanings. For example, according to the test publishers the correct heading for "A-bomb" is "Inventions." Yet the student who assigns "A-bomb" to the heading "Government" may have a far more sophisticated understanding of that concept than his test results would indicate.

One other point seems worth mentioning. On the surface, it appears that the same test task is required for both the Intermediate and Advanced levels above, that is, the assigning of vocabulary items to general headings. However, the difference in vocabulary items changes the skill required by the student considerably. Classifying famous people would seem to require less cognitive skill than classifying subjective terms such as "minority rights."

Test publishers have not clearly specified the cognitive skills assessed by standardized social studies tests. In many cases a variety of skills are lumped under an over-simplified heading such as "knowledge of social studies vocabulary." Test users must be cautious about judging a test by its name only.

4. Lack of an Adequate System for Developing Social Studies Test Items

Social studies test developers are suffering from a malady that has plagued reading test developers for a number of years. That is, lack of an adequate scheme or conceptual model for developing test items. To be sure, test publishers have become very sophisticated at data analysis after the items are constructed. Standardization procedures, item analyses, reliability and validity estimates, are conducted with efficiency and technical skill. Yet the actual development of test items is largely done on a logical and intuitive basis. The point is, our technical

expertise of what to do with items after they're developed far exceeds our knowledge of how to systematically construct items.

Consider for a moment the level of thinking regarding social studies test development. Most social studies tests are defined by the content they sample. References are made about the inclusion of American history, geography, sociology, or some other sub area of the social studies domain. Occasionally the tests, or portions of them, are defined in terms of student behavior such as the ability to read maps or interpret graphs.

One still sees recommendations in the literature to construct a topic-by-process matrix when designing a test. In such a matrix behaviors are crossed with elements of content. The behaviors are usually defined according to Bloom's Taxonomy (1956). This paradigm has been around for years and is still resorted to in a good many cases. The Taxonomy developed by Bloom and his associates sixteen years ago was a step in the right direction but it may have outlived its usefulness. As Bormuth (1970), Anderson (1972), Sullivan (1969), and many other critics have convincingly argued, the categories within the Taxonomy overlap and do not lead readily to operational definitions. Anyone who has tried to develop test items based on the Taxonomy can attest to the ambiguity involved. As Anderson (1972, p. 149) points out, "...what is required is a system of explicit definitions and rules to derive test items from instructional statements such that a person can answer the items correctly if, and only if, he comprehends the statements."

Do such Yes, there are such systems on the horizon, but at the present time they are exploratory and have not yet proven themselves. John Bormuth (1970), in his book, On the Theory of Achievement Test Items, presents a linguistic rationale for deriving test items from instructional statements. Essentially it involves making grammatical transformations to form different classes of test items.

Schlesinger and Weiser (1970) of the Israel Institute of Applied Social Research, have proposed a facet design for the systematic construction of items for a reading comprehension test. Their facet design, "...concentrates on the relationship between the test item and the text on which it is based, rather than on the skills and abilities presumably involved in answering the item." (p. 568) The classificatory scheme of this model would simultaneously include the correct answer as well as the incorrect distractors.

Another model that has received some attention is that of "domain-referenced achievement testing" proposed by Hively, et. al. (1968). In this approach, rules are specified to generate a universe or domain of every possible test item of interest in a field of knowledge. A test is formed by sampling from the universe in a partly random fashion. Hively and colleagues (1968) have worked out a system of rules for generating a universe of items to cover elementary mathematics. While elementary mathematics may be a relatively easy field in which to apply such a model, perhaps efforts should be made in other domains of knowledge, such as social studies.

As one can see, empirical item selection procedures are emerging. One of the common factors found in many of these new models is the relationship between the wording of the test item and the wording of instruction. Social studies test developers as well as social studies practitioners will need to become more linguistically-oriented in the future if they are to understand and apply rational item selection procedures.

5. Lack of Content Validity

This last major criticism could perhaps be considered a summary of all the criticisms made previously. Standardized social studies tests, for the most part, lack a clear and operational definition of content validity. It is not clear to test users what such tests are actually measuring. Teachers and principals use the test results to make statements about their students' levels of "social studies achievement" with only a vague and ambiguous understanding of that concept.

The American Psychological Association's Standards for Educational and Psychological Tests and Manuals (1966) makes the following recommendation regarding content validity:

"If a test performance is to be interpreted as a sample of performance or a definition of performance in some universe of situations, the manual should indicate clearly what universe is represented and how adequate is the sampling."

The concept of "social studies achievement" could be defined along two dimensions--the universe of content and the universe of behaviors. How adequately have publishers sampled from these two dimensions?

Content validity is frequently described in social studies test manuals as a process of carefully surveying and sampling textbooks, courses of study, or curriculum guides to obtain a representative segment of social studies content. However, rarely are the surveyed materials identified. Also, the copyright dates of the materials are almost never given. Critical reviews of social studies tests found in Buros (1971) suggest that the sampling has been less than adequate.

In terms of the universe of content, criticism has often been made of the overemphasis on history at the expense of multidisciplinary fields such as anthropology and sociology. Critics of the sampling from the universe of behaviors have pointed to emphasis on recall of factual information to the exclusion of such behaviors as critical reading, analysis, and application. Where test publishers have tried to define their test by both content and behaviors, the relationship between the two dimensions has been vague and not clearly spelled out.

Until test publishers adequately define the content of social studies tests, test users will have difficulty making meaningful interpretations of student performance.

Summary

The purpose of this paper was to critically examine standardized social studies achievement tests from a psychometric and reading perspective. Five major issues were identified that detract from the meaningful interpretation of student performance on standardized social studies tests. The issues discussed were (1) the reading-dependency of social studies items, (2) the picture-dependency of social studies items, (3) the cognitive skills assessed by social studies items, (4) the lack of an adequate system of item development, and (5) the lack of content validity. Each of the issues were defined and related to test validity and interpretation. The central theme running throughout the paper was the inadequacy of the content validity of most standardized social studies tests. Unless test publishers specify more explicitly the elements of content and types of behavior sampled on their tests, test users will have great difficulty making meaningful interpretations of student performance.

References

- American Psychological Association, Standards for Educational and Psychological Tests and Manuals, Washington, D. C. : American Psychological Association, 1966
- Anderson, R. C. "How to Construct Achievement Tests to Assess Comprehension," Review of Educational Research, 42(1972) 145-170
- Bloom, B. S. (ed.) Taxonomy of Educational Objectives: Cognitive Domain, New York: David McKay Inc., 1956
- Bormuth, J. R. On the Theory of Achievement Test Items, Chicago: University of Chicago Press, 1970
- Buros, O. K. The Seventh Mental Measurements Yearbook, volumes 1 and 2, Highland Park, New Jersey: Gryphon Press, 1972
- Gaines, W. G. An Application of John B. Carroll's Model of School Learning to the Teaching of Anthropology, unpublished doctoral dissertation, University of Georgia, 1971
- Gaines, W. G. and Jongsma, E. A. "The Effects of Instruction in Test-Taking Skills on Standardized Achievement Test Scores of Fifth Grade Pupils", in press, 1972
- Hively, W., Patterson, H. L. and Page, S. H. Generalization of Performance by Job Corp Trainees on a Universe-Defined System of Achievement Tests in Elementary Mathematical Calculation, Saint Paul: Minnesota National Laboratory, 1968
- Hively, W. "Domain-Referenced Achievement Testing", Symposium prepresented at the American Educational Research Association Convention, February, 1970
- Mitchell, R. W. "A Comparison of Children's Responses to an Original and Experimental Form of Subtests GS and ND of the Gates Basic Reading Test", unpublished doctoral dissertation, University of Minnesota, 1967
- Preston, R. C. "Ability of Students to Identify Correct Responses before Reading", Journal of Educational Research, 58 (1964), 181-183
- Preston, R. C. and Duffey, R. V. Primary Social Studies Test, Boston: Houghton Mifflin, 1967

- Schlesinger, I. M. and Weiser, Z. "A Facet Design for Tests of Reading Comprehension", Reading Research Quarterly, 5 (1970) 566-580
- Sullivan, H. J. "Objectives, Evaluation, and Improved Learner Achievement" in Instructional Objectives, R. E. Stake (ed.), American Educational Research Association Monograph Series on Curriculum Evaluation, Chicago: Rand McNally, 1969
- Thomas, G. The Use of Programmed Instruction for Teaching Anthropology in the Fifth Grade, unpublished doctoral dissertation, University of Georgia, 1967
- Tuinman, J. "Selected Aspects of the Assessment of the Acquisition of Information from Reading Passages", unpublished doctoral dissertation, University of Georgia, 1970
- Wash, J. A. An Evaluation of Pupil Performance in the Anthropology Curriculum Project, Grades 1, 2, 4, and 5, General Information Series No. 5, Anthropology Curriculum Project, University of Georgia, 1967
- Weaver, W. and Bickley, A. C. "Sources of Information for Responses to Reading Test Items," APA Proceedings, 75th Annual Convention, 1967, 293-294